

METHOD AND SYSTEM FOR FINDING SIMILAR RECORDS IN MIXED
FREE-TEXT AND STRUCTURED DATA

5

ABSTRACT OF THE DISCLOSURE

A technique for data mining where the available data contains both structured as well as unstructured (free-text) data. The present invention combines together the information available from different types of data to provide a single similarity score indicating the degree of similarity between records. Thus, a data evaluation application selects two records from a database and compares corresponding fields from the two records. The application determines whether to apply a nominal matching process, an ordinal matching process, or a vector-space matching process depending on the type of data in each pair of corresponding fields. The application sums the matching scores for all the fields in the records to compute the similarity score.